Research Theme: **Data Science**

MSc Research Project Title: **Signals and Noise in Academic Hiring: Modeling Career Outcomes at Scale**

Principal Investigator/Supervisor: **Richard She**

Co-supervisor/ Collaborator(s) (if any):

Project Description

a) Background:
Academic research is an engine of innovation—but the career trajectories of scientists are often governed by opaque, subjective processes. Who becomes a professor? Which credentials matter most? And are the selection criteria aligned with actual scientific impact?

Despite decades of discussion, we lack a quantitative, empirically grounded model of the academic job market. Hiring decisions are made by faculty search committees based on a mix of heuristics: publication records, institutional affiliations, fellowship prestige, letters of recommendation, and informal networks. But these inputs have never been systematically benchmarked. In effect, the scientific workforce is shaped by decisions that remain resistant to measurement—let alone accountability.

Our goal is to change that. This project seeks to model the predictors of academic hiring and success using large-scale data from public CVs, publication databases, and funding records. By assembling a linked dataset across training background, publication history, fellowship and grant support, and hiring outcomes, we aim to quantify which factors actually drive academic advancement—and which are reputational mirages. This work lies at the intersection of metascience, labor economics, and data science, with potential to inform both policy and individual career decisions.

b) Proposed work:
Incoming Ph.D. students will help build the first large-scale, data-driven framework for analyzing academic careers. This includes assembling structured datasets from faculty CVs, public grant databases, and bibliometric APIs (e.g. CrossRef, ORCID, NIH RePORTER), followed by model development to address key empirical questions in scientific labor markets.

Project 1: Predicting academic hiring using large-scale career trajectory data
We begin by modeling the outcomes of postdoctoral researchers on the academic job market. Core questions include: How predictive is obtaining a prestigious fellowship (e.g. HHWF, JCC, LSRF, Damon Runyon)? How much does a K99 award increase odds of hiring—and at what tier of institution? Do CNS publications matter more than consistent productivity in field-specific journals, and if so, by how much?

To answer these, we will apply multivariate regression and causal inference tools (e.g., instrumental variables, matching estimators) to a dataset spanning thousands of CVs and funding records. We will also explore ranking systems for institutions and departments, evaluate hiring outcomes by

demographic background over time, and measure the latent effect of advisor prestige or institutional affiliation using mixed models or hierarchical regressions.

A second phase of the project examines what happens after faculty are hired: How does research productivity change over time? Are there measurable patterns where professors become less scientifically active but take on more administrative roles, such as department chair or institute director? Can we build interactive dashboards that track research output over time, normalized by grant support or lab size?

A third branch of analysis focuses on lab ecosystems. How predictive is the success rate of a PI's trainees (e.g. fraction of postdocs who land faculty jobs)? Can we normalize lab output by number of trainees? What background traits (e.g. undergrad major, Ph.D. advisor, number of prior co-authors) distinguish high-impact PIs?

Students in this project will develop skills in data science, causal inference, natural language processing, and science policy analysis. They will help define new, transparent ways of measuring excellence in science—grounded in data, open to scrutiny, and designed to inform real decisions in funding, hiring, and mentorship.

c) Preferred skills:
Strong analytical skills and an interest in empirical modeling of social systems. Experience with Python, R, or data scraping tools is helpful but not required. No prior wet lab experience needed. Curiosity about the structure of science is essential.

Supervisor contact:

If you have questions regarding this project, please email the Principal Investigator:

richardsheshe@gmail.com

SBS contact and how to apply:

Associate Chair-Biological Sciences (Graduate Studies) : AC-SBS-GS@ntu.edu.sg

Please apply at the following:

Application portal:

https://venus.wis.ntu.edu.sg/GOAL/OnlineApplicationModule/frmOnlineApplication.ASPX